REMARKS

Claims 1-31, 35-44, 46-48, 52, and 53 are pending. Claims 1, 7, 9, 15, 17, 18, 35, 52, and 53 have been amended. No new matter has been entered. Claims 1-31, 35-44, 46-48, 52, and 53 remain.

The amendments present the rejected claims in better form for consideration on appeal and may be admitted pursuant to 37 C.F.R. § 1.116(b)(2).

An Information Disclosure Statement (IDS) is being submitted herewith. Acknowledgement of the IDS and entry of the cited art references on the record are requested.

Claims 1-23, 28, 29, 35-40, 46, 52, and 53 stand rejected under 35 U.S.C. § 103(a) as being obvious in light of International Application Publication No. WO 03/060766, to Lindh et al. ("Lindh"), in view of U.S. Patent No. 5,794,236, issued to Mehrle. Applicant traverses.

To establish a *prima facie* case of obviousness, the examiner has the burden of proving that (1) there is some suggestion or motivation, either in the references themselves or in the knowledge generally available to one of ordinary skill in the art, to modify the reference or combine the reference teachings; (2) there is a reasonable expectation of success; and (3) the combined references teach or suggest all the claim limitations. MPEP 2143. A *prima facie* of obviousness case has not been shown.

Lindh teaches pre-processing that involves extract all terms included in unformatted text and assigning weights to the terms based on their information content (*Lindh*, col. 16, lines 5-8). The information content can be determined by using an extension to a traditional Term Frequency times Inverse Document Frequency (TFIDF) term weighting scheme (*Lindh*, col. 17, line 20-col. 18, line 14). Lindh further teaches one method for enhancing relationship quality by filtering the document corpus used to generate a term-term matrix (*Lindh*, col. 27, lines 18-25). The reduction in the number of similar documents in the corpus can result in large quantities of similar documents not biasing the relationship measures, which is characterized as a flaw that can be reduced using document

clustering, such as *k*-means clustering (*Lindh*, col. 27, line 25-col. 28, line 5). A representative document vector is generated for each cluster found by means of a clustering algorithm, such as by calculating a cluster centroid as the mean of all document vectors in the cluster (*Lindh*, col. 28, lines 8-23). The representative document vector is added to the cluster and all other documents that belong to the cluster are removed from the initial document corpus (*Id.*).

Mehrle teaches a system and method for classifying legal documents into a hierarchy (*Mehrle*, col. 2, lines 31-40). Unique classification keys are mapped to each legal classification in the hierarchy to allow cumulative retrieval of all legal classifications occurring below any non-terminal node of the hierarchy (*Mehrle*, col. 6, lines 28-43). The legal hierarchy is annotated with seed citations and the number of levels and their contents will vary according to the particular subject matter (Col. 5, lines 38-62). A control file identifies the legal classifications, their classification keys, and the hierarchy location keys to which the classifications map (*Mehrle*, col. 7, lines 37-43). The control file is an intermediate file that is used as input to a legal classification generator along with a document to be classified (*Mehrle*, col. 8, lines 11-22). Citations from the unclassified document are compared to seed citations in the control file for matches and a classification score is incremented based on the of seed citations matched (*Mehrle*, col. 8, line 63-col. 9, line 10). The classification scores for the unclassified document are checked against a threshold after all citations have been checked (*Mehrle*, col. 9, lines 39-53 and col. 10, lines 4-5). If a classification score is equal to or exceeds the threshold, the classification key and hierarchy location key are inserted into the unclassified legal document (*Mehrle*, col. 9, lines 6-15).

Claims 1, 9, and 17 have been amended. Claim 1 now recites a threshold module determining similarities between the documents grouped into each cluster based on the center of the cluster and the scores assigned to each of the at least one concepts in each such document, dynamically determining a threshold for each cluster as a function of the similarities, and identifying and reassigning those documents having the similarities falling outside the threshold. Claim 9 now

recites determining similarities between the documents grouped into each cluster based on the center of the cluster and the scores assigned to each of the at least one concepts in each such document; dynamically determining a threshold for each cluster as a function of the similarities; and identifying and reassigning those documents having the similarities falling outside the threshold. Claim 17 now recites code for determining similarities between the documents grouped into each cluster based on the center of the cluster and the scores assigned to each of the at least one concepts in each such document; code for dynamically determining a threshold for each cluster as a function of the similarities; and code for identifying and reassigning those documents having the similarities falling outside the threshold. Support can be found in the specification on page 14, line 19-page 15, line 5 and page 20, line 10-page 21, line 13.

Claims 18, 35, 52, and 53 have also been amended. Claim 18 now recites a threshold module relocating outlier documents, comprising determining similarities between the documents grouped into each cluster based on the center of the cluster and the scores assigned to each of the at least one concepts in each such document, dynamically determining a threshold for each cluster as a function of the similarities, and identifying and reassigning the documents with the similarities falling outside the threshold. Claim 35 now recites relocating outlier documents, comprising determining similarities between the documents grouped into each cluster based on the center of the cluster and the scores assigned to each of the at least one concepts in each such document; dynamically determining a threshold for each cluster as a function of the similarities; and identifying and reassigning the documents with the similarities falling outside the threshold. Claim 52 now recites code for relocating outlier documents, comprising code for determining similarities between the documents grouped into each cluster based on the center of the cluster and the scores assigned to each of the at least one concepts in each such document; code for dynamically determining a threshold for each cluster as a function of the similarities; and code for identifying and reassigning the documents with the similarities falling outside the threshold.

Claim 53 now recites means for relocating outlier documents, comprising means for determining similarities between the documents grouped into each cluster based on the center of the cluster and the scores assigned to each of the at least one concepts in each such document; means for dynamically determining a threshold for each cluster as a function of the similarities; and means for identifying and reassigning the documents with the similarities falling outside the threshold. Support can be found in the specification on page 14, line 19-page 15, line 5 and page 20, line 10-page 21, line 13.

A *prima facie* case of obviousness has not been presented. First, Lindh teaches information retrieval for data organized in terms, which includes identifying clusters of documents with high similarity. Mehrle, on the other hand, teaches away from document clustering, specifically where the document being classified is used as a search source (*Mehrle*, col. 2, lines 1-28). Clustering legal documents clustered together on the basis of similarly occurring citations provides an unsatisfactory solution to automated legal classification, as a document is not assigned a classification in a legal hierarchy, which enables grouping under a common classification that exists independently from the actual citation. Thus, one of ordinary skill in the art would not find a suggestion or motivation to modify or combine the reference teachings of Lindh and Merle.

Second, Mehrle generates classification scores for an unclassified document by comparison to a control file, only after which the classification scores are compared to a threshold. The control file is an intermediate file that represents the legal hierarchy (*Mehrle*, col. 5, lines 11-17), wherein each seed citation in the control file signifies a point of law, that is, a proposition enunciated in a legal document supported by the legal authority from which the proposition is derived (*Mehrle*, col. 5, lines 1-4). Thus, the seed citations can be though of as a form of metadata, which is data about data, and any grouping of legal documents occurs based upon a particular seed citation indicates that legal documents in that grouping share the same point of law, yet do not necessarily contain the same literal content. One of ordinary skill in the art would not face a reasonable

expectation of success in combining the teachings of Lindh and Mehrle if that artisan were expecting Mehrle to provide an alternative to clustering, as the threshold taught by Mehrle operates to only admit, but not remove, legal documents into a grouping. Moreover, the threshold taught by Mehrle is for weighing classification scores that relate to data existing outside of the literal content of a legal document, that is, to a point of law. The use of the threshold taught by Mehrle would therefore result in an overly inclusive and overbroad criteria that would result in clusters that included documents that would not qualify for membership in a cluster if selected under a content-specific criteria.

Finally, the classification scores taught by Lindh-Mehrle signify those citations in an unclassified document that match seed citations in a control file. In contrast, the independent claims recite dynamically determining a threshold for each cluster as a function of the similarities, which are *directly* based upon the documents themselves, not seed citations in an intermediate file. Moreover, the threshold taught by Lindh-Mehrle fails to identify *and* reassign those documents having similarities falling outside the threshold, as recited by the independent claims. Thus, the combination of Lindh and Mehrle fail to teach or suggest all claim limitations.

Accordingly, a *prima facie* case of obviousness is not present for independent Claims 1, 9, 17, 18, 35, 52, and 53. Claims 2-8 are dependent on Claim 1 and are patentable for the above-stated reasons, and as further distinguished by the limitations therein. Claims 10-16 are dependent on Claim 9 and are patentable for the above-stated reasons, and as further distinguished by the limitations therein. Claims 19-23, 28, and 29 are dependent on Claim 18 and are patentable for the above-stated reasons, and as further distinguished by the limitations therein. Claims 36-40, 46 are dependent on Claim 35 and are patentable for the above-stated reasons, and as further distinguished by the limitations therein. Withdrawal of rejection is requested.

Claims 24-27 and 41-44 stand rejected under 35 U.S.C. § 103(a) as being obvious over Lindh as applied to Claims 18 and 35 above, and further in view of

U.S. Patent No. 6,675,159, issued to Lin et al. ("Lin"). Applicant traverses.

Claims 24-27 are dependent upon Claim 18 and are patentable for the reasons stated above with respect to the Lindh-Mehrle obviousness rejection, and as further distinguished by the limitations therein. Claims 41-44 are dependent upon Claim 35 and are patentable for the reasons stated above with respect to the Lindh-Mehrle obviousness rejection, and as further distinguished by the limitations therein. Withdrawal of rejection is requested.

Claims 30, 31, 47, and 48 stand rejected under 35 U.S.C. § 103(a) as being obvious over Lindh and Mehrle as applied to Claim 29 above, and further in view of Lin. Applicant traverses.

Claims 30 and 31 are dependent upon Claim 18 and are patentable for the reasons stated above with respect to the Lindh-Mehrle obviousness rejection, and as further distinguished by the limitations therein. Claims 47 and 48 are dependent upon Claim 35 and are patentable for the reasons stated above with respect to the Lindh-Mehrle obviousness rejection, and as further distinguished by the limitations therein. Withdrawal of rejection is requested.

Claims 1-31, 35-44, 46-48, 52, and 53 are believed to be in condition for allowance. Entry of the foregoing amendments is requested. Reconsideration of the claims, withdrawal of the finality of the Office action and a Notice of Allowance are earnestly solicited. Please contact the undersigned at (206) 381-3900 regarding any questions or concerns associated with the present matter.

Respectfully submitted,

Dated: February 12, 2007

By: _____
Patrick J.S. Inouye, Esq.
Reg. No. 40,297

Cascadia Intellectual Property
500 Union Street, Ste 1005
Seattle, WA 98101

Telephone: (206) 381-3900
Facsimile: (206) 381-3999